The genomic signature of trophic divergence along the benthic-limnetic axis in allopatric and sympatric threespine stickleback

Andreas Härer[1], Daniel Bonick[2], and Diana Rennison[1]

[1]University of California San Diego
[2]University of Connecticut

July 30, 2020

Abstract

The repeated occurrence of similar phenotypes in independent lineages (i.e., parallel evolution) in response to similar ecological conditions can provide compelling insights into the process of adaptive evolution. An intriguing question is to what extent repeated phenotypic changes are underlain by repeated changes at the genomic level and whether patterns of genomic divergence differ with the geographic context in which populations evolve. Here, we combine genomic, morphological and ecological datasets to investigate the genomic signatures of divergence across populations of threespine stickleback (Gasterosteus aculeatus) that adapted to contrasting trophic niches (benthic or limnetic) in either sympatry or allopatry. We found that genome-wide differentiation (FST) was an order of magnitude higher and substantially more repeatable for sympatric benthic and limnetic specialists compared to allopatric populations with similar levels of trophic divergence. We identified 55 genomic regions consistently differentiated between sympatric ecotypes that were also associated with benthic vs. limnetic niche across allopatric populations. These candidate regions were enriched on three chromosomes known to be involved in the benthic-limnetic divergence of threespine stickleback. Some candidate regions overlapped with QTL for body shape and trophic traits such as number of gill rakers, traits that strongly differ between benthic and limnetic ecotypes. In sum, our study shows that magnitude and repeatability of genomic signatures of trophic divergence in threespine stickleback highly depend on the geographical context. The identified candidate regions provide starting points to identify functionally important genes for the adaptation to benthic and limnetic trophic niches.

Keywords

parallel evolution, adaptation, genome-wide differentiation, *Gasterosteus aculeatus* , association mapping, trophic ecology

Introduction

The repeated evolution of similar phenotypes across independent lineages in response to shared environmental conditions (i.e., parallel evolution) provides strong evidence for natural selection (Manceau, Domingues, Linnen, Rosenblum, & Hoekstra, 2010; Rosenblum, Parent, Diepeveen, Noss, & Bi, 2017; Torres-Dowdall et al., 2017). Cases of parallel evolution have been described in a wide array of organisms across the tree of life (Colosimo et al., 2005; Mahler, Ingram, Revell, & Losos, 2013; Rosenblum et al., 2017; Sage, Christin, & Edwards, 2011). Parallel evolution was initially studied on the phenotypic level but recently focus has

shifted towards identifying examples on the molecular level (Stern, 2013). Phenotypic parallelism can be the product of mutations in the same gene (Chan et al., 2010; Rosenblum, Rompler, Schoneberg, & Hoekstra, 2010; Steiner, Rompler, Boettger, Schoneberg, & Hoekstra, 2009; Zhen, Aardema, Medina, Schumer, & Andolfatto, 2012) or involve many changes across the genome (Jones, Grabherr, et al., 2012; Ravinet et al., 2016; Rennison, Stuart, Bolnick, & Peichel, 2019) which results in a broad signature of parallel genomic divergence. Identifying examples of genetic and genomic parallelism improved our general understanding of parallel evolution (Arendt & Reznick, 2008; Manceau et al., 2010); repeated use of the same genes or genomic regions can suggest a source of genetic bias or constraint (reviewed in Bolnick, Barrett, Oke, Rennison, & Stuart, 2018) and the reuse of genes or regions can also be leveraged to identify candidate loci important for adaptation.

It is becoming clear that the magnitude of repeatability of genome-wide parallelism varies considerably across study systems (Jones, Chan, et al., 2012; Le Moan, Gagnaire, & Bonhomme, 2016; Ravinet et al., 2016). For example, species pairs of sunflowers that diverged along latitudinal gradients (Renaut, Owens, & Rieseberg, 2014) show high levels of genomic parallelism whereas little evidence for genomic parallelism is found in repeated adaptive radiations of Nicaraguan crater lake cichlid fishes (Kautt, Elmer, & Meyer, 2012). Within a species, population pairs can also vary in their magnitude of parallelism (e.g., Ravinet et al., 2016; Rennison et al., 2019). Threespine stickleback (*Gasterosteus aculeatus*) population pairs from adjacent lake and stream habitats in Canada show multiple highly divergent genomic regions. A substantial portion of these divergent regions (37%) is shared among independently evolved lake-stream pairs. In contrast, lake-stream pairs from Europe share only 3% of divergent regions (Feulner et al., 2015; Rennison et al., 2019). In the rough periwinkle (*Littorina saxatilis*), parallelism ranges from 8-34 % of outliers, depending on the populations compared (Kess, Galindo, & Boulding, 2018; Ravinet et al., 2016). Both spatial proximity and ecological similarity seem to be key predictors of the overall magnitude of genome-wide parallelism (Morales et al., 2019; Rennison, Delmore, Samuk, Owens, & Miller, 2020). A recent study on threespine stickleback from different areas of their global distribution further emphasized that the demographic history and previous selection can affect levels of genomic repeatability (Fang, Kemppainen, Momigliano, Feng, & Merila, 2020). Taken together, these results suggest that parallelism in genomic differentiation can be substantial but highly context dependent. Despite these research efforts, we currently lack a good understanding of how the geographic context (divergence in allopatry vs. sympatry) may affect patterns of genomic parallelism for populations adapting to similar ecological niches.

In sympatry, the lack of physical barriers allows for gene flow between diverging populations, which can counteract the accumulation of genome-wide differentiation (Coyne & Orr, 2004). Gene flow homogenizes neutral regions of the genome, and only few regions harboring genes under divergent selection are expected to be strongly differentiated when divergence occurs with gene flow, as shown for crows and *Heliconius* butterflies (Nadeau et al., 2014; Poelstra et al., 2014), although reinforcement could potentially mitigate this effect (Garner, Goulet, Farnitano, Molina-Henao, & Hopkins, 2018). The homogenizing effect of gene flow also reduces the fraction of the genome able to respond to natural selection (Samuk et al., 2017); previous work has shown that in the presence of gene flow, divergence is limited to regions with low rates of recombination (Samuk et al., 2017). Such constraints are not expected in allopatry and the stochastic effects of genetic drift, differences in effective population size and variable ecology may generate more inconsistent patterns of differentiation among allopatric populations. Thus, we predict higher levels of genomic parallelism across sympatric species due to the bias of divergence towards a smaller fraction of the genome and fewer stochastic peaks due to genetic drift.

Threespine stickleback represent an excellent system for studying the genomic signatures of repeated evolution in natural populations across different geographic settings. Stickleback have rapidly adapted to freshwater habitats throughout the northern hemisphere (Bell & Foster, 1994). Newly formed freshwater lakes were independently colonized by marine stickleback after the last ice age, around 10,000 - 12,000 years ago (Bell & Foster, 1994). Within these young lakes, stickleback have repeatedly and independently adapted to novel resources through parallel phenotypic evolution in trophic morphology (Bell & Foster, 1994; Bolnick & Ballare, 2020; Schluter & McPhail, 1992). Lakes vary in size and depth, encompassing different proportions

of benthic and limnetic habitat, which affects dietary and habitat availability for stickleback (Bolnick & Ballare, 2020). Accordingly, variation in diet and morphology across allopatric populations is associated with lake size; stickleback mostly feed on littoral invertebrates (benthic prey) in small lakes and pelagic zooplankton (limnetic prey) in large lakes (Bolnick & Ballare, 2020). In medium-sized lakes, stickleback generally have intermediate phenotypes and broader dietary niches (Bolnick & Ballare, 2020). While most lakes are inhabited by a solitary population (morphologically unimodal for most traits and approximately panmictic), in five lakes in British Columbia the colonizing stickleback independently evolved into co-occurring pairs of sympatric benthic and limnetic specialists (Taylor & McPhail, 1999). This repeated divergence in trophic ecology along the benthic-limnetic axis across sympatric and allopatric stickleback populations allows us to study parallelism of genomic differentiation in different geographic settings.

Here, we employ two approaches to map the genomic signatures of sticklebacks' adaptation to benthic and limnetic habitats. We use F_{ST} to detect adaptive divergence between benthic and limnetic sympatric species pairs (Gow, Rogers, Jackson, & Schluter, 2008; Schluter & McPhail, 1992) and among allopatric populations from small benthic and large limnetic lakes (Bolnick & Ballare, 2020). Further, we use genome-wide association (GWA) mapping (Bolnick & Ballare, 2020) for a larger dataset of allopatric lake populations to detect alleles associated with lake size (the proxy for dietary niche). By comparing benthic-limnetic adaptation in different geographic contexts, we were able to quantify the magnitude of parallelism and ask whether the geographic context affects patterns of shared genomic architecture during adaptation to similar niches. Furthermore, it is likely that regions identified to overlap between these datasets contain loci important for adaptation to divergent benthic and limnetic niches, such candidate regions provide opportunities for follow-up work.

Materials and Methods

Data acquisition

We reanalyzed previously collected morphological, ecological and genomic data of wild-caught stickleback from British Columbia, Canada, obtained from several independent datasets. The genotype data for the genome-wide association (GWA) mapping of lake size was generated using ddRAD sequencing (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012) for 33 solitary stickleback populations from Vancouver Island. For 21 of these solitary populations, we obtained data on the proportion of benthic diet and gill raker numbers (a key trophic trait). Both of these datasets are from Bolnick and Ballare (2020); the genomic dataset was produced by Stuart et al. (2017). Genotype data used for F_{ST}analyses from three sympatric benthic-limnetic species pairs (Paxton, Priest and Little Quarry Lakes) were generated by Samuk et al. (2017) using the genotyping-by-sequencing method. Diet and gill raker number data for the sympatric species pairs from Paxton and Priest Lakes were obtained from Schluter and McPhail (1992). QTL data on different sympatric species pairs were obtained from several previous studies (Arnegard et al., 2014; Conte et al., 2015; Malek, Boughman, Dworkin, & Peichel, 2012).

Morphological analyses

Linear regression models were used to test whether lake size (log surface area) affects trophic ecology (proportion of benthic diet) or morphology (gill raker number), as previously shown for allopatric populations (Bolnick & Ballare, 2020). Linear regression was done twice, once with only the 21 solitary populations and once including the sympatric benthic-limnetic species pairs. We further determined residuals to identify populations that deviate strongest from the linear regression models (Fig. S1). The proportion of benthic diet consumed by stickleback has been shown to be affected by lake size across allopatric populations (Bolnick & Ballare, 2020). In the lakes harboring sympatric benthic-limnetic species pairs, that show pronounced specialization in diet, the association between lake size and diet is assumed to be broken up leading to extreme residuals for some of these sympatric species (Fig. S1).

Genome-wide differentiation analyses and quantification of parallelism

To investigate the genomic architecture of divergence along the benthic-limnetic axis, we determined patterns of genome-wide differentiation using Weir and Cockerham's F_{ST} (Weir & Cockerham, 1984) for eight solitary populations and three benthic-limnetic species pairs. The eight solitary populations were selected to reflect strong differences in trophic ecology and included populations from the four smallest (Little Goose, Little Mud, Muskeg, Ormund) and the three largest (Lower Campbell, Upper Campbell, Stella) lakes. Additionally, the population from Amor Lake was included in the group of large lakes (although it is only the sixth largest lake) because it had the lowest proportion of benthic diet (8.9%). To allow comparisons across datasets, we calculated mean F_{ST}values for 50-kb windows for windows with a minimum of three data points and all analyses were based on these windows. A 50-kb window was classified as an outlier if the F_{ST} value was in the top 5% of the genome-wide F_{ST} distribution. Spearman's rank correlation coefficients were used to quantify the extent of shared genome-wide differentiation across pairs of populations that diverged along the benthic-limnetic axis.

We sought to detect genomic regions that are repeatedly differentiated either across benthic-limnetic species pairs or across solitary populations from small and large lakes. Levels of repeatability were estimated as in Rennison et al. (2019). Briefly, repeatability was calculated for each window as the proportion of population comparisons with data for which this window was scored as an F_{ST}outlier. Statistical significance of repeatability for the solitary comparisons was determined at the 0.05 level by comparing empirical values against a null distribution produced by 10,000 permutations with subsampling to account for non-independence of some comparisons. P-values were corrected for multiple testing using the Benjamini & Hochberg method (Benjamini & Hochberg, 1995).

To examine the overlap between F_{ST} outliers for sympatric species pairs and lake size GWA mapping candidates (see next paragraph for more information) for the 33 solitary populations, we normalized benthic-limnetic F_{ST} values and took the mean across species pairs. As we were interested in F_{ST}outlier regions *consistently* associated with benthic-limnetic differentiation, only F_{ST} windows with data for at least two species pairs were used. Windowed F_{ST} values were normalized using the following equation:

$$x - \text{mean}(x)/\text{SD}(x),$$

where x is a vector of all windowed F_{ST} values for each population pair. Normalized F_{ST} values for each window were then averaged across the three lakes.

Genome-wide association mapping

The GWA mapping for lake size (the proxy for dietary niche) was based on 175,350 single nucleotide polymorphisms (SNPs) obtained from ddRAD sequencing of 33 solitary populations with 12 fish per population, as detailed in Bolnick and Ballare (2020). For more information on library preparation and bioinformatics pipeline protocols, see Stuart et al. (Stuart et al., 2017). To test for associations, a binominal linear model with allele frequency and log lake size was run for each SNP with a logit link function and watershed as a covariate, see Bolnick and Ballare (2020) for further details. The p-values resulting from this analysis were used to identify candidate genomic regions (50-kb windows) associated with lake size; a window was considered a GWA candidate if it contained loci with a p-value falling below the 0.05 cutoff.

Overlap of candidate regions between datasets

All analyses that involved comparisons of F_{ST} and GWA mapping data were based on 4,985 50-kb windows distributed across all 21 chromosomes that were included in both datasets. We created four subsets of the combined benthic-limnetic F_{ST} and GWA mapping dataset. These subsets contained windows that were either non-significant in both datasets, F_{ST} outliers in the sympatric benthic-limnetic data (95^{th} percentile), significantly associated with lake size across solitary lake populations ($P < 0.05$) or double outliers in which they were both an F_{ST} outlier and significantly associated with lake size. While identifying windows that

were F_{ST} outliers and significantly associated with lake size allowed us to obtain a set of candidate regions that might be important during adaptation to benthic and limnetic habitats, we want to point out the differences between these two analyses. The GWA mapping allowed us to detect specific alleles significantly associated with lake size, it provides information on the directionality of shifts in allele frequencies across many populations. In contrast, the F_{ST} analysis lacks such directionality and rather tells us whether the same genomic loci are repeatedly differentiated but not which alleles are more common in which populations. Thus, these are independent lines of evidence suggesting a given genomic region contributes to adaptation along the benthic-limnetic axis.

We tested whether windows of these subsets (except for non-significant windows) were significantly enriched on certain chromosomes. This was done by running a permutation with 10,000 iterations. For each iteration, the number of significant windows in each subset were randomly shuffled across different chromosomes (accounting for chromosome size). Empirical frequencies of significant windows for each chromosome were then compared to the null distribution obtained by permutation and significant enrichment was determined at the 0.05 level. Chi-squared tests of independence were used to determine whether there was a significant association between F_{ST} outliers and GWA candidate windows. Spearman's rank correlation coefficients were calculated to test whether the GWA p-values are correlated with normalized benthic-limnetic F_{ST} values. To obtain further information on the phenotypic traits associated with candidate regions, we identified published quantitative trait loci (QTL) whose peaks overlapped with windows differentiated along the benthic-limnetic axis. All statistical analyses were performed in R v3.5.1 (R Core Team, 2016).

Results

Phenotypic and ecological variation along the benthic-limnetic axis

Across the 21 solitary stickleback populations, trophic ecology (proportion of benthic food items; Fig. 1A) and morphology (gill raker number; Fig. 1B) varied with lake size (log surface area; Table S1). Stickleback from smaller lakes preferentially consumed benthic prey (linear regression, $R^2 = 0.6117$, $P < 0.001$) and had more gill rakers ($R^2 = 0.3039$, $P = 0.006$). Sympatric benthic-limnetic species pairs from Paxton and Priest Lake fell towards the ends of these distributions. Benthic species mainly consumed benthic prey and had fewer gill rakers whereas limnetic species consumed limnetic prey and had more gill rakers (Fig. 1). Including benthic-limnetic species pairs into linear regression models substantially reduced the proportion of variation explained by the models for diet ($R^2 = 0.3038$, $P = 0.003$) and gill raker number ($R^2 = 0.1415$, $P = 0.036$). This can be explained by the fact that the residuals of the sympatric species pairs were consistently among the highest (Fig. S1), and specialization appears independent of lake size.

Genomic architecture of benthic-limnetic differentiation in sympatry and allopatry

Mean genome-wide F_{ST} between allopatric solitary populations from small and large lakes were much lower (0.026 - 0.059; Fig. S2) than between benthic and limnetic species occurring in sympatry (0.199 - 0.226; Fig. S3). These differences remained when F_{ST} outliers were removed (solitary populations: 0.015 - 0.042; sympatric species pairs: 0.166 - 0.193). Genome-wide patterns of differentiation were largely shared across sympatric benthic-limnetic species pairs but not among solitary populations or between solitary populations and benthic-limnetic species pairs; correlation coefficients for F_{ST} values across the genome were low for the small-large lake solitary population pairs (mean $\rho = 0.023$, 'solitary' in Fig. 2) and much higher for the benthic-limnetic species pairs (mean $\rho = 0.524$, 'benthic-limnetic' in Fig. 2). When comparing genome-wide patterns in F_{ST} between the small-large lake solitary population pairs and the benthic-limnetic species pairs, correlation coefficients were very low (mean $\rho = -0.002$, 'solitary vs. benthic-limnetic' in Fig. 2).

Across benthic-limnetic species pairs, we identified 287 F_{ST} outlier windows based on mean normalized F_{ST}

values. These windows were unequally distributed, with some chromosomes showing significant enrichment ($P< 0.05$, chromosomes 1, 4, 7 and 20) and others completely lacking outliers (chromosomes 3, 6, 10, 14 and 15; Fig. 3A). Chromosome 7, which constitutes around 7% of the genome, had the highest proportion of outliers (14%; Fig. S4A). For the sympatric benthic-limnetic comparison, a total of 67 windows were repeatedly differentiated (outlier in at least two of the three population pairs), representing 23.3% of F_{ST} outlier windows. These windows were spread across more than half of the chromosomes (12 out of 21; Fig. 3A and Fig. S3). In the allopatric solitary comparison for small and large lakes, 520 windows were significantly repeatedly differentiated ($P < 0.05$ after correction for multiple testing), and these were spread across all chromosomes (Fig. S2). Seven of these allopatric-divergence windows were also repeatedly differentiated in the benthic-limnetic species pairs, which were located on chromosomes 4,7, 18 and 19.

To identify candidate regions differentiated along the benthic-limnetic axis across all 33 solitary populations, we also used genome-wide association (GWA) mapping. This approach afforded more power by leveraging the variance of more populations. Using this approach, we found 1,015 of the 4,985 windows surveyed (20.4%) to be significantly associated with lake size (our proxy for dietary niche; $P< 0.05$), although none of these windows remained significant after very conservative Bonferroni correction. The proportion of significant windows (20.4%) was much higher than would be expected when assuming a 5% false positive rate and we were interested in the overlap between datasets, which would hint at biological significance. Hence, we proceeded in analyzing the candidate loci classified as significant before Bonferroni correction. These candidate windows were distributed across all 21 chromosomes (Fig. 3B), but significantly enriched ($P < 0.05$) on chromosomes 11, 16 and 19 (Fig. 3B and Fig. S4B).

A comparison of our candidates from the lake size GWA mapping in solitary populations and F_{ST} outliers from the sympatric benthic-limnetic species pairs led to the identification of genomic regions likely to be important for adaptation to contrasting benthic and limnetic niches in different geographic settings. There were 55 candidate windows shared between the two datasets (hereafter referred to as 'double outliers'), which represents 1.1% of the total 4,985 windows (Fig. 4). Double outliers were enriched on chromosomes 4, 7 and 19 (Fig. S4C). A chi-squared test of independence suggested that this overlap between the two datasets was not more than expected by chance; candidate windows significantly associated with lake size were not more likely to be also F_{ST} outliers, χ^2 (1, N = 4,985) = 0.336, $P = 0.562$. There was a weak, but significant, negative correlation between normalized F_{ST} values and association mapping -$\log_{10}(P)$ values across all 4,985 shared windows (Spearman's rank correlation, ρ = -0.049, $P< 0.001$). Performing correlation tests for different categories (non-significant, F_{ST} outliers, significantly associated with lake size, double outliers; Fig. 4) showed a weak negative correlation for non-significant windows (ρ = -0.079, $P< 0.001$). Spearman's rank correlation coefficients were positive (but not statistically significant) for the significantly associated with lake size and double outlier categories (Fig. S5).

Characterization of candidate regions

In total, 13 previously described QTL from different sympatric species pairs (Arnegard et al., 2014; Conte et al., 2015; Malek et al., 2012) overlapped with benthic-limnetic F_{ST} outlier windows. These QTL containing F_{ST} outliers were mostly located on chromosomes 1, 4 and 7 and included body shape and trophic traits such as number of gill rakers (Table S2). An additional 16 QTL included windows significantly associated with lake size, but not with sympatric divergence. These QTL included body shape, dorsal spine length, number of lateral plates and gill rakers and suction feeding index (Table S2). Three QTL contained double outlier windows and these were associated with color and shape of males as well as body shape (Table S2). Notably, most QTL (10 out of 13) were located on chromosomes enriched for F_{ST} outlier windows but no such pattern was observed for significant lake size candidate windows (1 out of 16). For double outlier, one out of three QTL mapped to chromosome 7, which was significantly enriched (Table S2, Fig. 4C).

Discussion

Geographic context of divergence and parallelism

While sympatric benthic and limnetic species as well as allopatric solitary populations showed substantial variation in morphology and diet (Matthews, Marchinko, Bolnick, & Mazumder, 2010; Schluter & McPhail, 1992), the magnitude and patterns of genomic differentiation of benthic and limnetic ecotypes differed with geographic context (sympatry vs. allopatry). Sympatric benthic and limnetic species pairs show strong evidence of reproductive isolation (Rundle, Nagel, Wenrick Boughman, & Schluter, 2000). In line with these observations, we found that genomic differentiation between species was high in all three lakes (Fig. S3). Despite the reduced opportunity for gene flow afforded by allopatry, genomic differentiation between solitary populations from small and large lakes was an order of magnitude lower than that found for the sympatric species pairs. The strong divergence of benthic and limnetic pairs in sympatry suggests that these ecotypes may be further along the speciation continuum; perhaps these populations are moving towards a genome-wide phase of divergence, sometimes called 'genome-wide congealing' (Feder et al., 2014). In contrast, low levels of genomic divergence as observed for the solitary populations may indicate a more 'genic phase' of adaptation, with divergence limited to fewer loci (Feder et al., 2014). Linked selection (both genetic/genome hitchhiking and background selection) may also contribute to the observed higher degree of differentiation in the sympatric comparisons (Flaxman, Feder, & Nosil, 2013; Nosil, Funk, & Ortiz-Barrientos, 2009). These disparate patterns could be influenced by the process of reinforcement, where selection against hybrids increases reproductive isolation between lineages occurring in sympatry (e.g., Noor, 1999). Accordingly, a previous study on benthic and limnetic stickleback revealed female preference for mates of the same ecotype only in sympatric populations but not in allopatric ones (Rundle & Schluter, 1998).

Patterns of genome-wide differentiation were very consistent across sympatric species pairs (Fig. 2 and Fig. S3), and 23.3% of outlier regions were repeatedly differentiated. This is remarkable given that these species pairs have evolved independently in different lakes within the last 12,000 years (Taylor & McPhail, 1999). A study by Jones et al. surveying the same three lakes showed similar levels of repeatability, but was based on a lower number of genomic loci (Jones, Chan, et al., 2012). The repeated divergence of the same genomic regions for these sympatric species pairs suggests that a genetic constraint or bias may favor certain variants, loci, and/or regions during adaptation (reviewed in Bolnick et al., 2018). Given the young age of these species, standing genetic variation is expected to be more important for adaptation rather than *de novo* mutations. In stickleback, adaptation from standing genetic variation is thought to have been important for rapid adaptation to freshwater habitats (Colosimo et al., 2005; Jones, Chan, et al., 2012). The observed high degree of parallelism for the sympatric species pairs may indicate that adaptation from standing genetic variation has also been important for the repeated adaptation to benthic and limnetic niches (Jones, Chan, et al., 2012). Previous work on the sympatric benthic-limnetic pairs has also shown that divergence is heavily biased towards regions of the genome with suppressed recombination (Samuk et al., 2017); this bias may contribute to the similarity of genome-wide patterns, as only adaptive loci found within regions of low recombination may be able to diverge substantially.

The solitary populations and sympatric species pairs differ to a similar extent in trophic ecology and morphology (Fig. 1), yet the patterns of genomic parallelism are remarkably different. In contrast to the strong evidence of genome-wide parallelism found for the sympatric species pairs, there was no evidence that the same genomic regions were consistently differentiated (i.e., high F_{ST}) across solitary populations from different lakes exhibiting benthic-limnetic divergence (Fig. 2 and Fig. S2). Stickleback lake-stream pairs from Canada and Europe show similarly low levels of parallelism (Feulner et al., 2015; Rennison et al., 2019). The absence of parallelism for these solitary populations potentially means that standing genetic variation is not central to benthic-limnetic divergence in allopatry or that genetic variation may be more limited in some populations (Stuart et al., 2017), perhaps due to bottlenecks during colonization of upstream habitats. Also, it is possible that the marine ancestors were more heterogeneous than typically assumed (e.g., Stuart et al., 2017) and so different watersheds may have had genetically distinct founders. Population-specific

effects of gene flow and genetic drift may also contribute to low levels of observed parallelism (Fitzpatrick, Torres-Dowdall, Reznick, Ghalambor, & Funk, 2014; Stuart et al., 2017). It would be interesting to have estimates of the demographic history of the stickleback populations investigated here, e.g., on the size of the founder populations. Genetic drift is stronger in small populations, which will reduce levels of parallelism across populations (MacPherson & Nuismer, 2017; Szendro, Franke, de Visser, & Krug, 2013). Further, if founder populations were small, there is also a high chance that the genetic diversity that selection could act upon was distinct across lakes, presumably limiting parallelism in genome-wide patterns of differentiation. Heterogeneity of the recombination landscape also doesn't seem to play an important role when populations are diverging in allopatry (Samuk et al., 2017), which might contribute to different signatures of adaptation between the allopatric and sympatric comparisons.

Differences in the selective landscape among populations may also explain low levels of parallelism among solitary population pairs and between sympatric and allopatric comparisons. Theoretical work suggests that genomic parallelism decreases rapidly as the selection landscape becomes less parallel (Thompson, Osmond, & Schluter, 2019). In Trinidadian guppies, a famous example for parallel evolution, life-history traits and mortality were non-parallel between low and high predation, which could be attributed to stream-specific variation in disease and flooding (Fitzpatrick et al., 2014). Parallelism in morphological divergence of stream-lake pairs of stickleback decreases as environmental differences become non-parallel (Stuart et al., 2017). In marine-freshwater pairs of stickleback and in the ecotypes of the rough periwinkle, parallelism is also dependent on ecological similarity (Morales et al., 2019; Rennison et al., 2020). Even though all populations were adapting to similar benthic and limnetic niches, there may be cryptic environmental heterogeneity across the different lakes and within the discrete habitat categories that reduces overall parallelism in the selective landscape.

Identification of candidate regions for benthic-limnetic adaptation

Identification of repeatedly diverged genomic regions can increase our understanding of the predictability of evolutionary trajectories (Conte, Arnegard, Peichel, & Schluter, 2012; Stern & Orgogozo, 2008) and can help to detect important adaptive genes or traits as genetic drift is unlikely to result in patterns of repeated divergence (Schluter & Nagel, 1995). A handful of chromosomes appear to contribute disproportionately in the adaptation to benthic and limnetic niches. When looking at the genomic locations of candidate regions obtained from multiple analyses within this study, we detected evidence for significant clustering on chromosomes 4, 7 and 19. These chromosomes harbor multiple windows that were repeatedly differentiated across both sympatric benthic-limnetic species pairs *and* solitary populations from small and large lakes that differ substantially in their trophic ecology. Windows that were highly differentiated across sympatric benthic-limnetic species pairs *and* significantly associated with lake size across allopatric populations (double outliers) were also significantly enriched on chromosomes 4, 7 and 19. These chromosomes appear to be hot spots in threespine stickleback evolution (Conte et al., 2015; Jones, Chan, et al., 2012), and are enriched for multiple categories of ecologically relevant QTL (Peichel & Marques, 2017). Future comparative work examining the structure and evolutionary history of these chromosomes may provide insights into why these chromosomes are so central to the benthic-limnetic adaptation in this system.

To infer the functional relevance of genomic regions related to benthic-limnetic divergence, we incorporated published QTL data (Arnegard et al., 2014; Conte et al., 2015; Malek et al., 2012). We defined candidate regions as genomic windows that were either F_{ST} outliers across sympatric benthic-limnetic species pairs, significantly associated with lake size across allopatric populations or falling in both categories. The QTL that mapped to candidate regions were primarily located on chromosomes 4 and 7 (12 out of 32), again suggesting that these chromosomes are key to benthic-limnetic adaptation. Most QTL affected general body shape and five QTL affected the number of gill rakers (Table S2), a trait that has been shown to be important for the trophic ecology of these fish (Schluter, 1993). Benthic fish that feed on relatively large littoral invertebrates have fewer and shorter gill rakers than limnetic fish that feed on smaller pelagic zooplankton, both in sympatric and allopatric populations (Fig. 1; Bell & Foster, 1994; Schluter, 1993).

Two more QTL affected other aspects of the trophic apparatus and feeding performance (Table S2). Note, however, that entire categories of potentially adaptive phenotypes have yet to be subjected to QTL analysis of benthic-limnetic divergence (e.g., immune traits, metabolic traits, the gut microbiota, etc.).

By integrating different types of data and looking for repeatability across sympatric and allopatric comparisons, we were able to identify 55 strong candidate regions underlying adaptation to the contrasting benthic and limnetic niches, several of which contain QTL for important morphological traits. These 55 candidate regions were F_{ST} outliers in benthic-limnetic species pairs*and* associated with lake size across solitary populations; hence, they are presumably targets of divergent selection. It is unlikely that the same regions were identified by chance as we used independent datasets, methods with different biases and had a fairly strict criterion for inclusion. This suggests that our candidate regions merit further molecular analyses to identify the genetic variants underlying adaptation to benthic and limnetic niches. Such investigation would reveal to what extent the parallelism in genomic regions observed here represents parallel changes in the same genes or perhaps even the same mutations. Parallelism is predicted to be more prevalent at higher levels of biological organization, as shown in the rough periwinkle (Ravinet et al., 2016) and Australian groundsel (Roda et al., 2013) were patterns of parallelism increased from SNPs to genomic regions to molecular pathways. Fine-mapping of genetic variants within these regions may provide good candidates for future gene editing studies in order to isolate their phenotypic and fitness effects.

Caveats

One limitation of our study is that we assessed parallelism at the level of genomic regions, which means that we cannot say whether the same loci or alleles underlie the observed patterns of differentiation. A strength of our approach is the integration of multiple datasets and types of data. However, a drawback of this approach is that we integrate sequence datasets generated using different methods (Genotyping-by-Sequencing vs. ddRAD sequencing). This means that some genomic regions were not surveyed in both datasets, so it is possible that we have underestimated the amount of parallelism between the sympatric and allopatric comparisons. Additionally, the use of arbitrary significance cutoffs means that we likely missed some regions that are evolving in parallel.

Conclusion

Our assessment of genomic parallelism in allopatry vs. sympatry revealed substantial differences in the magnitude of genomic divergence during adaptation to benthic and limnetic niches under different geographic contexts. Integration of independent datasets showed that parallelism is genome-wide among the three extant sympatric species pairs but very limited among solitary populations. By comparing candidate loci from the sympatric species pairs with those identified in solitary populations, we were able to identify candidate regions that might be essential for threespine sticklebacks' ecological divergence along the benthic-limnetic axis.

Acknowledgements

We thank members of the threespine stickleback research community, whose body of empirical work we built upon. This work was supported by funding from the University of California San Diego to D.J.R. and funding from NIH 1R01AI123659-01A1 and NSF DEB-1144773 grants to D.I.B.

References

Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution, 23* (1), 26-32. doi:10.1016/j.tree.2007.09.011

Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S., . . . Schluter, D. (2014). Genetics of ecological divergence during speciation. *Nature, 511* (7509), 307-311. doi:10.1038/nature13301

Bell, M. A., & Foster, S. A. (1994). *The Evolutionary Biology of the Threespine Stickleback* , Oxford, UK: Oxford University Press.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing.*ournal of the Royal Statistical Society: Series B (Statistical Methodology), 57* (1), 289-300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bolnick, D. I., & Ballare, K. M. (2020). Resource diversity promotes among-individual diet variation, but not genomic diversity, in lake stickleback. *Ecology Letters, 23* (3). doi:10.1111/ele.13448

Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non)parallel evolution. *Annual Review of Ecology, Evolution, and Systematics, 49* , 303-330. doi:10.1146/annurev-ecolsys-110617-062240

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Jr., Shapiro, M. D., Brady, S. D., . . . Kingsley, D. M. (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science, 327* (5963), 302-305. doi:10.1126/science.1182213

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Jr., Dickson, M., Grimwood, J., . . . Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science, 307* (5717), 1928-1933. doi:10.1126/science.1107239

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., . . . Peichel, C. L. (2015). Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. *Genetics, 201* (3), 1189-1200. doi:10.1534/genetics.115.182550

Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B-Biological Sciences, 279* (1749), 5039-5047. doi:10.1098/rspb.2012.2146

Coyne, J. A., & Orr, H. A. (2004). *Speciation* , Sunderland, MA: Sinauer Associates.

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merila, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution* . doi:10.1038/s41559-020-1222-6

Feder, J. L., Nosil, P., Wacholder, A. C., Egan, S. P., Berlocher, S. H., & Flaxman, S. M. (2014). Genome-wide congealing and rapid transitions across the speciation continuum during speciation with gene flow. *Journal of Heredity, 105* , 810-820. doi:10.1093/jhered/esu038

Feulner, P. G. D., Chain, F. J. J., Panchal, M., Huang, Y., Eizaguirre, C., Kalbe, M., . . . Milinski, M. (2015). Genomics of divergence along a continuum of parapatric population differentiation. *PLoS Genetics, 11* (2). doi:10.1371/journal.pgen.1004966

Fitzpatrick, S. W., Torres-Dowdall, J., Reznick, D. N., Ghalambor, C. K., & Funk, W. C. (2014). Parallelism isn't perfect: could disease and flooding drive a life-history anomaly in Trinidadian guppies? *American Naturalist, 183* (2), 290-300. doi:10.1086/674611

Flaxman, S. M., Feder, J. L., & Nosil, P. (2013). Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution, 67* (9), 2577-2591. doi:10.1111/evo.12055

Garner, A. G., Goulet, B. E., Farnitano, M. C., Molina-Henao, Y. F., & Hopkins, R. (2018). Genomic signatures of reinforcement. *Genes, 9* (4), 191. doi:10.3390/genes9040191

Gow, J. L., Rogers, S. M., Jackson, M., & Schluter, D. (2008). Ecological predictions lead to the discovery of a benthic-limnetic sympatric species pair of threespine stickleback in Little Quarry Lake, British Columbia. *Canadian Journal of Zoology, 86* (6), 564-571. doi:10.1139/Z08-032

Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., . . . Kingsley, D. M. (2012). A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology, 22* (1), 83-90. doi:10.1016/j.cub.2011.11.045

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., . . . Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature, 484* (7392), 55-61. doi:10.1038/nature10944

Kautt, A. F., Elmer, K. R., & Meyer, A. (2012). Genomic signatures of divergent selection and speciation patterns in a 'natural experiment', the young parallel radiations of Nicaraguan crater lake cichlid fishes.*Molecular Ecology, 21* (19), 4770-4786. doi:10.1111/j.1365-294X.2012.05738.x

Kess, T., Galindo, J., & Boulding, E. G. (2018). Genomic divergence between Spanish Littorina saxatilis ecotypes unravels limited admixture and extensive parallelism associated with population history.*Ecology and Evolution, 8* (16), 8311-8327. doi:10.1002/ece3.4304

Le Moan, A., Gagnaire, P. A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact.*Molecular Ecology, 25* (13), 3187-3202. doi:10.1111/mec.13627

MacPherson, A., & Nuismer, S. L. (2017). The probability of parallel genetic evolution from standing genetic variation. *Journal of Evolutionary Biology, 30* (2), 326-337. doi:10.1111/jeb.13006

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science, 341* (6143), 292-295. doi:10.1126/science.1232392

Malek, T. B., Boughman, J. W., Dworkin, I., & Peichel, C. L. (2012). Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Molecular Ecology, 21* (21), 5265-5279. doi:10.1111/j.1365-294X.2012.05660.x

Manceau, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B., & Hoekstra, H. E. (2010). Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society B-Biological Sciences, 365* (1552), 2439-2450. doi:10.1098/rstb.2010.0104

Matthews, B., Marchinko, K. B., Bolnick, D. I., & Mazumder, A. (2010). Specialization of trophic position and habitat use by sticklebacks in an adaptive radiation. *Ecology, 91* (4), 1025-1034. doi:10.1890/09-0235.1

Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin, R. K. (2019). Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast.*Science Advances, 5* (12). doi:10.1126/sciadv.aav9963

Nadeau, N. J., Ruiz, M., Salazar, P., Counterman, B., Medina, J. A., Ortiz-Zuazaga, H., . . . Papa, R. (2014). Population genomics of parallel hybrid zones in the mimetic butterflies, H. melpomene and H. erato. *Genome Research, 24* (8), 1316-1333. doi:10.1101/gr.169292.113

Noor, M. A. F. (1999). Reinforcement and other consequences of sympatry.*Heredity, 83* , 503-508. doi:10.1038/sj.hdy.6886320

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology, 18* (3), 375-402. doi:10.1111/j.1365-294X.2008.03946.x

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks.*Philosophical Transactions of the Royal Society B-Biological Sciences, 372* (1713), 20150486. doi:10.1098/rstb.2015.0486

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One, 7* (5), e37135. doi:10.1371/journal.pone.0037135

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Muller, I., . . . Wolf, J. B. W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science, 344* (6190), 1410-1414. doi:10.1126/science.1253226

R Core Team. (2016). *R: A language and environment for statistical computing* , Vienna, Austria. Retrieved from https://www.R-project.org/.

Ravinet, M., Westram, A., Johannesson, K., Butlin, R., Andre, C., & Panova, M. (2016). Shared and nonshared genomic divergence in parallel ecotypes of Littorina saxatilis at a local scale. *Molecular Ecology, 25* (1), 287-305. doi:10.1111/mec.13332

Renaut, S., Owens, G. L., & Rieseberg, L. H. (2014). Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology, 23* (2), 311-324. doi:10.1111/mec.12600

Rennison, D. J., Delmore, K. E., Samuk, K., Owens, G. L., & Miller, S. E. (2020). Shared patterns of genome-wide differentiation are more strongly predicted by geography than by ecology. *American Naturalist, 195* (2), 192-200. doi:10.1086/706476

Rennison, D. J., Stuart, Y. E., Bolnick, D. I., & Peichel, C. L. (2019). Ecological factors and morphological traits are associated with repeated genomic differentiation between lake and stream stickleback.*Philosophical Transactions of the Royal Society B-Biological Sciences, 374* (1777), 20180241. doi:10.1098/rstb.2018.0241

Roda, F., Liu, H. L., Wilkinson, M. J., Walter, G. M., James, M. E., Bernal, D. M., . . . Ortiz-Barrientos, D. (2013). Convergence and divergence during the adaptation to similar environments by an Australian groundsel. *Evolution, 67* (9), 2515-2529. doi:10.1111/evo.12136

Rosenblum, E. B., Parent, C. E., Diepeveen, E. T., Noss, C., & Bi, K. (2017). Convergent phenotypic evolution despite contrasting demographic histories in the fauna of White Sands. *American Naturalist, 190* (S1), S45-S56. doi:10.1086/692138

Rosenblum, E. B., Rompler, H., Schoneberg, T., & Hoekstra, H. E. (2010). Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences of the United States of America, 107* (5), 2113-2117. doi:10.1073/pnas.0911042107

Rundle, H. D., Nagel, L., Wenrick Boughman, J., & Schluter, D. (2000). Natural selection and parallel speciation in sympatric sticklebacks.*Science, 287* (5451), 306-308. doi:10.1126/science.287.5451.306

Rundle, H. D., & Schluter, D. (1998). Reinforcement of stickleback mate preferences: Sympatry breeds contempt. *Evolution, 52* (1), 200-208. doi:10.2307/2410935

Sage, R. F., Christin, P. A., & Edwards, E. J. (2011). The C(4) plant lineages of planet Earth. *Journal of Experimental Botany, 62* (9), 3155-3169. doi:10.1093/jxb/err048

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology, 26* (17), 4378-4390. doi:10.1111/mec.14226

Schluter, D. (1993). Adaptive radiation in sticklebacks - size, shape, and habitat use efficiency. *Ecology, 74* (3), 699-709. doi:10.2307/1940797

Schluter, D., & McPhail, J. D. (1992). Ecological character displacement and speciation in sticklebacks. *American Naturalist, 140* (1), 85-108. doi:10.1086/285404

Schluter, D., & Nagel, L. M. (1995). Parallel speciation by natural selection. *American Naturalist, 146* (2), 292-301. doi:10.1086/285799

Steiner, C. C., Rompler, H., Boettger, L. M., Schoneberg, T., & Hoekstra, H. E. (2009). The genetic basis of phenotypic convergence in Beach Mice: Similar pigment patterns but different genes.*Molecular Biology and Evolution, 26* (1), 35-45. doi:10.1093/molbev/msn218

Stern, D. L. (2013). The genetic causes of convergent evolution.*Nature Reviews Genetics, 14* (11), 751-764. doi:10.1038/nrg3483

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: How predictable is genetic evolution? *Evolution, 62* (9), 2155-2177. doi:10.1111/j.1558-5646.2008.00450.x

Stuart, Y. E., Veen, T., Weber, J. N., Hanson, D., Ravinet, M., Lohman, B. K., . . . Bolnick, D. I. (2017). Contrasting effects of environment and genetics generate a continuum of parallel evolution. *Nature Ecology & Evolution, 1* , 0158. doi:10.1038/s41559-017-0158

Szendro, I. G., Franke, J., de Visser, J. A. G. M., & Krug, J. (2013). Predictability of evolution depends nonmonotonically on population size.*Proceedings of the National Academy of Sciences of the United States of America, 110* (2), 571-576. doi:10.1073/pnas.1213613110

Taylor, E. B., & McPhail, J. D. (1999). Evolutionary history of an adaptive radiation in species pairs of threespine sticklebacks (Gasterosteus): insights from mitochondrial DNA. *Biological Journal of the Linnean Society, 66* (3), 271-291. doi:10.1006/bijl.1998.0266

Thompson, K. A., Osmond, M. M., & Schluter, D. (2019). Parallel genetic evolution and speciation from standing variation. *Evolution Letters, 3* (2), 129-141. doi:10.1002/evl3.106

Torres-Dowdall, J., Pierotti, M. E. R., Härer, A., Karagic, N., Woltering, J. M., Henning, F., . . . Meyer, A. (2017). Rapid and parallel adaptive evolution of the visual system of Neotropical Midas cichlid fishes. *Molecular Biology and Evolution, 34* (10), 2469-2485. doi:10.1093/molbev/msx143

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution, 38* (6), 1358-1370. doi:10.1111/j.1558-5646.1984.tb05657.x

Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M., & Andolfatto, P. (2012). Parallel molecular evolution in an herbivore community.*Science, 337* (6102), 1634-1637. doi:10.1126/science.1226630

Data Accessibility

Sequence reads for the ddRAD dataset used for GWA mapping and F_{ST} analyses of solitary populations were archived under http://web.corral.tacc.utexas.edu/Stuart_2017_NatureEE_Data_Code/ on the University of Texas Austin Corral server by Stuart et al. (2017). Sequence reads for the genotyping-by-sequencing dataset used for F_{ST} analyses of benthic-limnetic species pairs were deposited on the NCBI Sequence Read Archive under the project number SRP107890 (see Table S1 in Samuk et al., 2017 for more information). The R code used for analyses and the underlying data files will be archived in the Dryad database upon acceptance.

Author Contributions

D.J.R. and D.I.B. developed the project; A.H. conducted the analyses and wrote the manuscript with input from D.J.R. and D.I.B. All of the authors approved the final version of the manuscript.

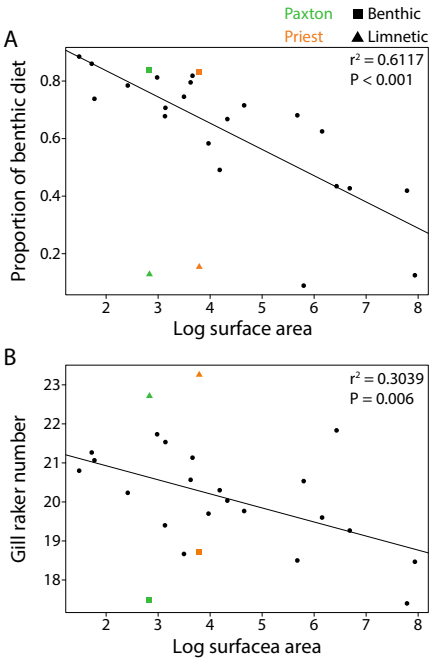

Figure 1: Linear regression models show an decrease in benthic diet (A) and gill raker number (B) with lake size for 21 solitary stickleback populations (black points). Sympatric species pairs of benthic and limnetic stickleback are not included in the regression models and are indicated in colored symbols. Each data point represents the population mean (see Table S1 for sample sizes).

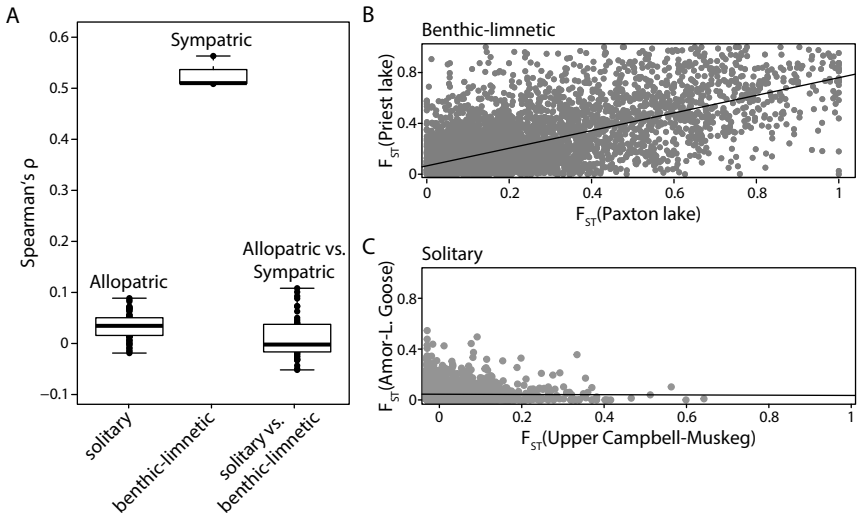

Figure 2: Spearman's rank correlation coefficients of pairwise comparisons of genome-wide F_{ST} values (A). Correlation coefficients were calculated between solitary populations with the lowest and highest proportions of benthic diet (allopatric, as exemplified in C), benthic-limnetic species pairs (sympatric, as exemplified in B) and among solitary populations and benthic-limnetic species pairs (allopatric vs. sympatric). All pairwise F_{ST} comparisons between solitary populations contained four different lakes (e.g., Amor-Little Goose vs. Lower Campbell-Muskeg).

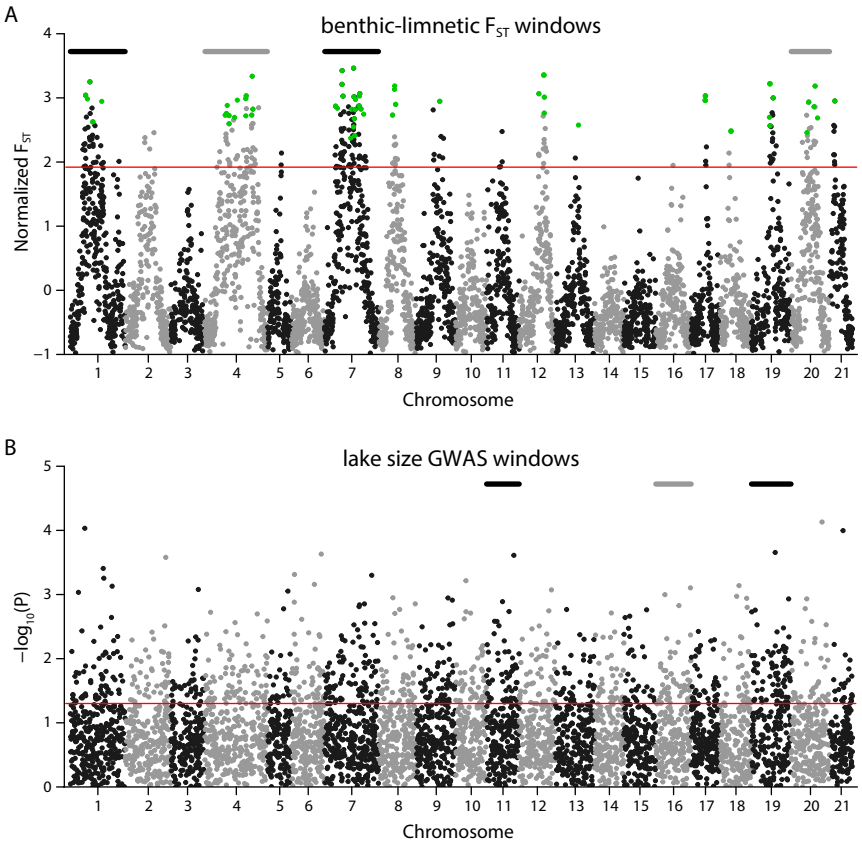

Figure 3: Normalized F_{ST} between benthic-limnetic species pairs from three lakes (A) and genome-wide association mapping for lake size across 33 solitary stickleback populations (B). Data points represent 50-kb windows and either show average F_{ST} (A) or the lowest p-value (B) for each window. Red lines indicate the 95th percentile for benthic-limnetic F_{ST} and the significance cutoff for the lake size GWAS ($P < 0.05$). Green data points represent repeated F_{ST} outliers for the benthic-limnetic species pairs. Horizontal black and grey lines highlight chromosomes with significant enrichment of F_{ST} outlier windows (A) or significant lake size GWAS windows (B).

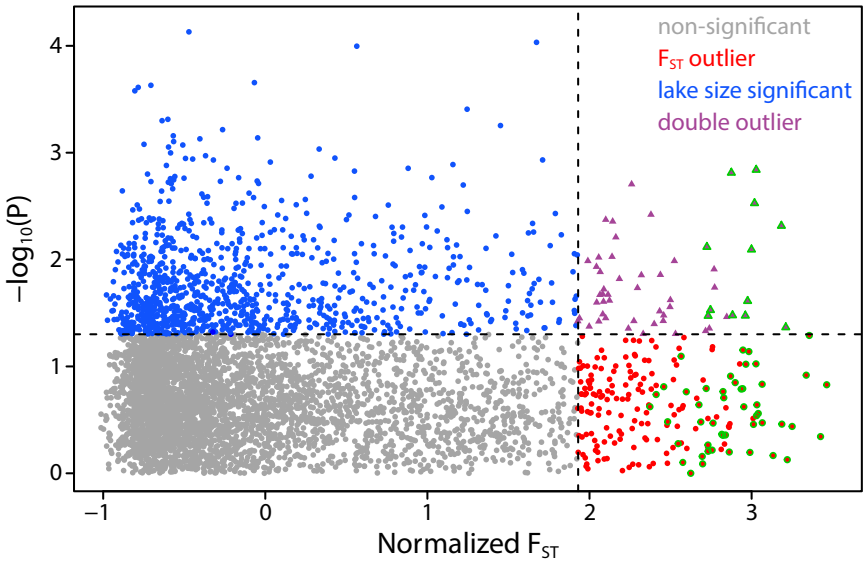

Figure 4: Scatterplot of GWAS -$\log_{10}(P)$ values for lake size and normalized F_{ST} values between benthic and limnetic species. 50-kb windows falling in different categories (non-significant, F_{ST} outlier, GWAS significant, double outlier) are indicated in colors and separated by dashed lines. Data points with green edges represent repeated F_{ST} outliers for the three benthic-limnetic species pairs (see also Figure 3A).